[CANCER RESEARCH 63, 3872–3876, July 15, 2003]

Advances in Brief

Gene Expression Analysis of Esophageal Squamous Cell Carcinoma Reveals Consistent Molecular Profiles Related to a Family History of Upper Gastrointestinal Cancer

Hua Su,[1] Nan Hu,[1] Joanna Shih, Ying Hu, Quan-Hong Wang, Eric Y. Chuang, Mark J. Roth, Chaoyu Wang, Alisa M. Goldstein, Ti Ding, Sanford M. Dawsey, Carol Giffen, Michael R. Emmert-Buck, and Philip R. Taylor[2]

Cancer Prevention Studies Branch [H. S., N. H., M. J. R., C. W., S. M. D., P. R. T.], Biometric Research Branch [J. S.], Laboratory of Population Genetics [Y. H.], Radiation Oncology Branch [E. Y. C.], Genetic Epidemiology Branch [A. M. G.], and Laboratory of Pathology [M. R. E-B.], National Cancer Institute, NIH, Bethesda, Maryland 20892; Shanxi Cancer Hospital and Institute, Taiyuan, 030013, People's Republic of China [Q-H. W., T. D.]; and Information Management Service, Inc., Silver Spring, Maryland 20904 [C. G.]

Abstract

Tumor and matched normal tissue from 19 esophageal squamous cell carcinoma patients from a high-risk area of China were analyzed with 7680 gene cDNA microarrays. Forty-one genes were differentially expressed ($P < 0.001$; ≥2-fold change) between tumor and matched normal samples (13 overexpressed and 28 underexpressed). Hierarchical clustering showed consistent molecular profiles across patients. Multidimensional scaling plots visually distinguished cases by family history status, which was confirmed statistically using a global permutation test ($P = 0.007$); we then identified 152 genes of which the expression differed in tumors from family history positive *versus* negative cases (55 overexpressed and 97 underexpressed at $P < 0.001$). These data indicate that molecular profiles in esophageal squamous cell carcinoma are highly consistent and that expression patterns in familial cases differ from those in sporadic cases.

Introduction

ESCC[3] is one of the most common fatal cancers worldwide, and Shanxi Province, a region in north-central China, has some of the highest rates in the world (1). Although epidemiological studies indicate that tobacco smoking and alcohol consumption are the major risk factors for squamous esophageal cancer in the low-risk regions of Europe and North America, the etiological agents in high-risk regions have yet to be convincingly identified. Within these high-risk regions, studies have shown a strong tendency toward familial aggregation, suggesting that genetic susceptibility, in conjunction with potential environmental exposures, may be involved in the etiology of this cancer (2–5). To better understand the role of genetics in the etiology of ESCC and to identify potential susceptibility genes, we previously conducted genome-wide studies of allelic loss in ESCC patients from Shanxi Province. Results suggested there is an overall high level of genetic instability in ESCCs from this area, and additionally indicated that chromosome 13 may harbor a tumor susceptibility gene in the population (6–8).

To complement the genomic studies of ESCC, we are also analyzing tumor gene expression profiles. Array technologies are comprehensive and relatively accurate ways to simultaneously analyze the expression of thousands of genes, and these technologies have been used to clarify gene expression changes in many human malignancies, including nine published studies of squamous cell carcinoma and/or adenocarcinoma of the esophagus (9–17). However, at least half of the microarray expression studies of ESCC published to date have relied on cell culture systems (9, 10, 12, 14). Cell lines differ from tumor cells, as they have been removed from their *in vivo* environment and are selected for growth characteristics in culture, thus bringing into question the clinical relevance of these findings. In this study, tumor and matched normal tissue from 19 ESCC patients were analyzed using cDNA microarrays containing 7680 genes to evaluate gene expression differences in ESCC patients from a high-risk area in China.

Materials and Methods

Patient Selection

Patients presenting in 2000 and 2001 to the Shanxi Cancer Hospital in Taiyuan, Shanxi Province, People's Republic of China, who were diagnosed with ESCC and considered candidates for curative surgical resection were identified and recruited to participate in this study. The study was approved by the Institutional Review Boards of the Shanxi Cancer Hospital and the NCI. None of the patients had prior therapy, and Shanxi was the ancestral home for all. After obtaining informed consent, patients were interviewed to obtain information on demographic and lifestyle cancer risk factors, and clinical data were collected. Tumor tissue and matching normal tissue distant to the tumor were obtained during surgery, snap-frozen in liquid nitrogen, and stored at −130 C until used. Specimens were chosen for this study based on two criteria: (*a*) histological diagnosis of ESCC confirmed by pathologists at both the Shanxi Cancer Hospital (Q-H. W.) and the NCI (M. J. R.); and (*b*) sufficiently high purity of tissue (the percentage of tumor for our specimens ranged from 50 to 95%, with a median of 85%) from both neoplastic cells in the tumor and non-neoplastic cells in the matched normal tissue.

Microarray Fabrication

The microarrays used for the experiments (NCI ROSP 8K Human Array)[4] contained 7680 human cDNA clones and were prepared from the Research Genetics Named Genes set (Huntsville, AL). These cDNA clones are all known genes and can be classified into several groups based on their biological functions, such as stress proteins, cell cycle control proteins, signal transduction proteins, apoptosis, transcription factors, DNA repair and replication proteins, cytokines, and so forth. All of the 7.6K cDNAs were spotted onto poly-L-lysine-coated slides using an OmniGrid arrayer (GeneMachines, San Carlos, CA) according to Eisen and Brown (18).

Sample Preparation and Chip Hybridization

Total RNA Isolation. Total RNA was extracted from frozen tumor and matched normal tissue by using TRIzol reagent (Invitrogen, Carlsbad, CA)

Received 3/10/03; accepted 5/22/03.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] These authors contributed equally to this work.

[2] To whom requests for reprints should be addressed, at Cancer Prevention Studies Branch, National Cancer Institute, 6116 Executive Boulevard, Room 705, Bethesda, MD 20892-8314. Phone: (301) 592-2932; Fax: (301) 435-8645; E-mail: ptaylor@mail.nih.gov.

[3] The abbreviations used are: ESCC, esophageal squamous cell carcinoma; NCI, National Cancer Institute; MDS, multidimensional scaling; UGI, upper gastrointestinal.

[4] Internet address: http://nciarray.nci.nih.gov/.

Table 1 *Demographic and clinical characteristics of ESCC cases*

No.	Patient ID	Age/sex	Tumor location	Pathological grade/ tumor stage	# lymph nodes positive/ # lymph nodes examined	Family history of any cancer
1	SHE1190	56/F	Middle	G3/3	2/6 (Y)	2 esophageal cancer (paternal-uncles), liver cancer (father)
2	SHE1447	63/M	Middle	G2/3	1/5 (Y)	Esophageal cancer (sister), rectal cancer (father)
3	SHE1451	59/M	Middle	G2/3	0/15 (N)	2 esophageal cancer (father and brother)
4	SHE1478	46/F	Middle	G3/3	0/14 (N)	Cardia cancer (father), liver cancer (brother)
5	SHE1493	64/F	Upper	G2/3	1/12 (Y)	Cardia cancer (brother)
6	SHE1496	50/M	Middle	G2/3	0/7 (N)	Esophageal cancer (father)
7	SHE0948	62/F	Middle	G3/3	2/5 (Y)	N
8	SHE1184	44/F	Middle	G1/3	0/8 (N)	Larynx cancer (mother), uterus cancer (sister)
9	SHE1187	40/M	Middle	G3/3	5/14 (Y)	N
10	SHE1188	58/F	Middle	G2/3	2/11 (Y)	Lung cancer (paternal-uncle)
11	SHE1201	64/M	Middle	G2/3	0/11 (N)	N
12	SHE1203	57/M	Middle	G2/3	0/12 (N)	N
13	SHE1330	65/F	Middle	G2/3	0/11 (N)	N
14	SHE1409	54/M	Middle	G2/3	8/12 (Y)	Pancreas cancer (paternal-grandfather)
15	SHE1410	57/M	Middle	G2/3	8/18 (Y)	N
16	SHE1418	60/F	Middle	G2/3	1/19 (Y)	N
17	SHE1479	47/F	Middle	G2/3	0/6 (N)	N
18	SHE1488	66/M	Middle	G2/3	0/13 (N)	Larynx cancer (brother)
19	SHE1489	59/F	Middle	G2/3	3/25 (Y)	N

Table 2 *Summary of 41 genes over- or underexpressed in ESCC*

No.	Gene/location on chromosome	Putative function	UniGene	Mean[a]	SD[b]	*P*	No. of cases	Fold change
Overexpressed								
1	*Collagen type III α 1 (COL3A1)*/2q31	Extracellular matrix	Hs.119571	1.54	1.00	<0.0001	19	2.91
2	*Keratin 14 (KRT14)*/17q12–q21	Cytoskeleton	Hs.117729	1.38	1.40	0.0004	19	2.60
3	*Collagen, type VII, α (COL7A1)*/3p21.1	Extracellular matrix	Hs.1640	1.21	0.95	<0.0001	19	2.31
4	*SNL (Fascin)*/7p22	Cytoskeleton/regulation of cellular processes	Hs.118400	1.15	0.82	<0.0001	19	2.22
5	*SPARC*/5q31.3–q32	Extracellular matrix remodelling	Hs.111779	1.06	0.73	<0.0001	17	2.08
6	*Transgelin 2 (TAGLN2)*/1q21–q25	Cytoplasmic/anti-pathogen response	Hs.75725	1.04	0.78	<0.0001	19	2.06
7	*Laminin γ 2 (LAMC2)*/1q25–q31	Extracellular matrix	Hs.54451	1.04	0.97	0.0002	19	2.06
8	*FADD=MORT*/11q13.3	Death receptor/cytoplasmic response	Hs.86131	1.03	0.95	0.0006	16	2.04
9	*Cystatin SN (CST1)*/20p11.21	Cysteine protease inhibitor	Hs.123114	1.03	1.05	0.0005	19	2.04
10	*HLA-B*/6p21.3	Immune system	Hs.77961	1.02	1.00	0.0003	19	2.03
11	*SCYB10*/4q21	Positive control of cell proliferation/cell motility/extracellular matrix	Hs.2248	1.02	1.06	0.0005	19	2.03
12	*CDC25B*/20p13	Mitosis/cell cycle control	Hs.153752	1.01	1.00	0.0003	19	2.01
13	*Collagen, type I, alpha 2 (COL1A2)*/7q22.1	Extracellular matrix	Hs.179573	1.01	0.75	<0.0001	19	2.01
Underexpressed								
14	*Keratin 4 (KRT4)*/12q12–q13	Cytoskeleton	Hs.3235	−2.67	2.34	0.0001	19	0.16
15	*Transglutaminase 3 (TGM3)*/20q11.2	Protein modification	Hs.2022	−2.28	1.62	<0.0001	17	0.21
16	*Cystatin A (stefin A) (CSTA)*/3q21	Proteinase inhibitor	Hs.2621	−1.94	1.74	0.0001	19	0.26
17	*NICE-1 protein (NICE-1)*/1q21	Epidermal differentiation	Hs.110196	−1.90	1.78	0.0002	19	0.27
18	*Small proline-rich protein SPRK (SPRK)*/1q21–q22	Cross-linked envelope protein of keratinocytes	Hs.46320	−1.86	1.94	0.0006	19	0.28
19	*FOS-like antigen 1(FOSL1)*/11q13	Positive control of cell proliferation/ cellular defense response	Hs.283565	−1.81	1.49	0.0001	19	0.29
20	*Uroplakin 1A (UPK1A)*/19q13.1	Cell-surface protein/regulation of cell development/activation/growth/motility	Hs.159309	−1.77	1.48	0.0001	19	0.29
21	*Epithelial membrane protein 1 (EMP1)*/12p12.3	Cell death/oncogenesis/cell proliferation	Hs.79368	−1.70	1.53	0.0001	19	0.31
22	*SPINK5*/5q32	Serine protease inhibitor	Hs.331555	−1.69	1.46	0.0001	19	0.31
23	*Cystatin B (stefin B) (CSTB)*/21q22.3	Protease inhibitor	Hs.695	−1.57	1.56	0.0003	19	0.34
24	*Adipose specific 2 (APM2)*/10q23.2	Unknown	Hs.74120	−1.55	1.14	<0.0001	19	0.34
25	*Proline rich 4 (lacrimal) (PROL4)*/12p13	Extracellular processes	Hs.45033	−1.52	1.51	0.0003	19	0.35
26	*ARS component B (ARS)*/8q24.3	Cell activation/cell adhesion	Hs.103505	−1.48	1.03	<0.0001	19	0.36
27	*Kallikrein 13 (KLK13)*/19q13.3–q13.4	Serine protease	Hs.165296	−1.42	1.48	0.0006	19	0.37
28	*Calponin 3, acidic (CNN3)*/1p22–p21	Cytoskeleton	Hs.194662	−1.41	1.10	<0.0001	19	0.38
29	*S100A9*/1q21	Calcium binding/signal transduction/ extracellular space	Hs.112405	−1.40	1.46	0.0005	19	0.38
30	*Betacellulin (BTC)*/4q13–q21	Positive control of cell proliferation	Hs.73105	−1.40	1.30	0.0002	19	0.38
31	*Hypothetical protein EDAG-1 (EDAG-1)*/9q22.32	Transforming gene, associated with development of tumor	Hs.176626	−1.34	1.29	0.0003	19	0.40
32	*Periplakin (PPL)*/16p13.3	Cytoskeleton	Hs.74304	−1.28	1.38	0.0008	19	0.41
33	*Early growth response 1 (EGR1)*/5q31.1	Transcriptional regulator/DNA binding	Hs.326035	−1.21	1.33	0.0009	19	0.43
34	*Adenomatous polyposis coli like (APCL)*/19p13.3	β-Catenin binding/tumor suppressor/ signal transduction	Hs.20912	−1.18	1.30	0.0009	19	0.44
35	*Kallikrein 11 (KLK11)*/19q13.3–q13.4	Serine-type peptidase	Hs.57771	−1.14	1.23	0.0008	19	0.45
36	*HPGD*/4q34–q35	Inactivation of prostaglandins/enzyme/ electron transporter	Hs.77348	−1.14	1.26	0.001	19	0.45
37	*ASE-1*/19q13.3	rDNA transcription	Hs.211956	−1.08	1.13	0.0006	19	0.47
38	*DUSP5=dual specificity phosphatase 5*/10q25	Heat shock response/protein dephosphorylation	Hs.2128	−1.08	1.01	0.0002	19	0.47
39	*RAR-beta=retinoic acid receptor, β (BARB)*/3p24	Epidermal banding/retinoic acid receptor/oncogenesis	Hs.171495	−1.07	1.07	0.0004	19	0.48
40	*Enoplakin (EVPL)*/17q25	Epidermal differentiation and maintenance	Hs.25482	−1.07	0.89	0.0001	19	0.48
41	*Extracellular matrix protein 1 (ECM)*/1q21	Extracellular matrix	Hs.81071	−1.01	0.98	0.0003	19	0.50

[a] Mean is the mean of log-ratio (base 2).
[b] SD is the standard deviation of log-ratio (base 2).

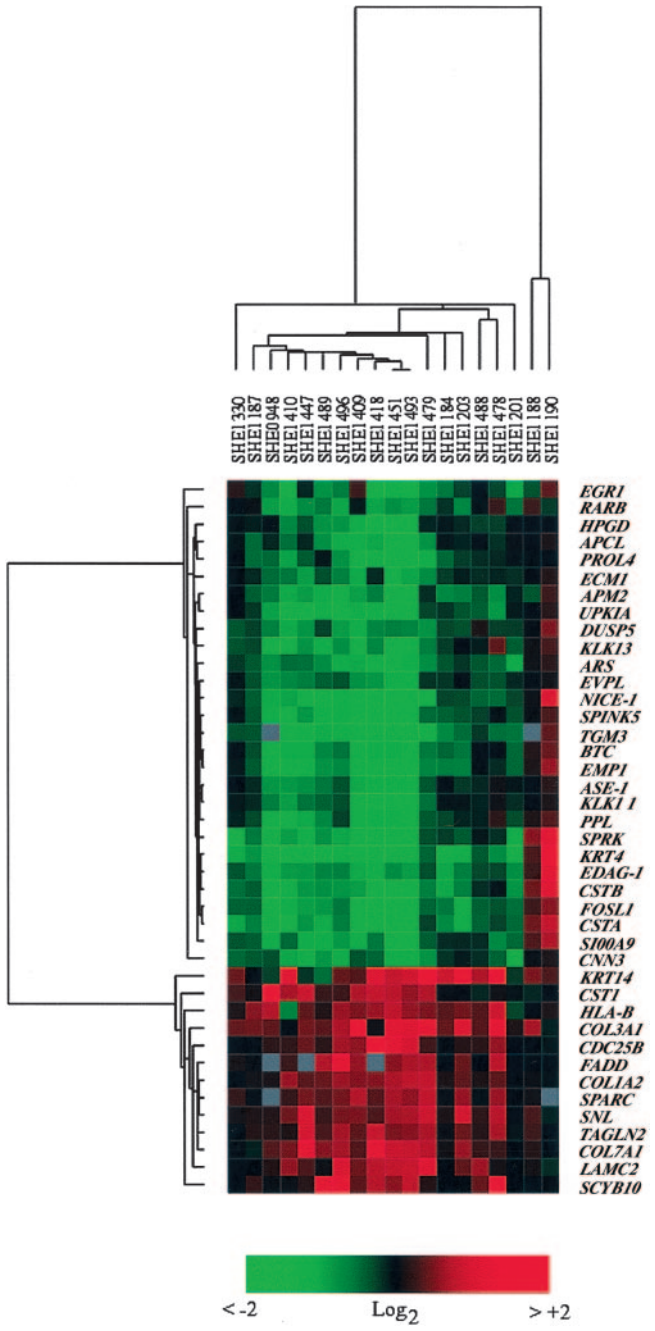

Fig. 1. Hierarchical clustering of the 41 genes that were over- or underexpressed in tumor *versus* normal tissue in the 19 ESCC patients.

following the protocol of the manufacturer. The integrity of total RNA was checked on 1.2% denaturing agarose gel electrophoresis (visual presence of 28S and 18S bands).

mRNA Amplification. One to 3 μg total RNA was used in each mRNA amplification. Antisense RNAs were generated by two rounds of amplification as described previously by Wang *et al.* (19).

Labeling and Hybridization in the Dye Swap Design. Three μg of antisense RNA from the tumor tissue and the matched normal tissue were labeled with Cy3-dUTP and Cy5-dUTP (Amersham Pharmacia Biotech, Piscataway, NJ), respectively, and called the forward group. Reciprocal labeling was also performed for each case (*i.e.,* tumor labeled with Cy5-dUTP and normal tissue-labeled with Cy3-dUTP) and called the reverse group. The labeled probes for the forward group were hybridized with the 7.6K chip in parallel with that of the reverse group in a hybridization buffer for 14–16 h at 65°C. After hybridization, the slides were washed in 2× SCC with 0.1% SDS, 1× SCC, 0.2× SCC, and 0.05× SCC, sequentially for 1 min each, and then spin dried.

Microarray Image Analysis

Hybridized arrays were scanned at 10 μm resolution on a GenePix 4000A scanner (Axon Instruments, Inc., Foster City, CA) at variable PMT voltage settings to obtain maximal signal intensities with <0.1% probe saturation. The Cy5-labeled cDNAs were scanned at 635 nm, and the Cy3-labeled cDNA samples were scanned at 532 nm. The resulting TIFF images were analyzed by GenePix Pro 3.0 software (Axon Instruments, Inc.). Both digitized images were overlaid to form a pseudo-colored image, and a detection method was then used to determine the actual target region based on the information from both red (Cy5) and green (Cy3) pixel values. The ratios of the sample intensity to the reference intensity (red: green) for all of the targets were determined, and ratio normalization was performed to normalize the center of the ratio distribution to 1.0. The intensities of each hybridization signal were evaluated by the NCI microarray database.[4]

Data Analysis

Missing Data. Log_2-ratios of median local background subtracted intensity levels were analyzed. Those array elements with intensity <100 in one channel were truncated at 100, and those array elements that were flagged as bad spots during image analysis or that had intensities <100 in both channels were treated as missing. Three percent of the array elements were found to be missing in this way. Those patients with missing values for a particular array element were excluded from all of the analyses involving that array element.

Normalization and Adjustment for Dye Bias. Log_2-ratios for each microarray were normalized by the locally robust smoother Lowess (20). Residual dye bias after normalization was removed by taking the average of forward (green-tumor/red-normal) and reverse (red-tumor/green-normal) fluorescent log_2-ratios for each of the samples where both forward and reverse labeling were done. The spot-specific dye bias was estimated from these samples by taking half of the average difference between forward and reverse log_2-ratios. For some samples, reverse labeling was not successful, so the estimate of the dye bias was subtracted from the log_2-ratio.

Generation of the Gene List. One-sample *t* tests were used to identify genes that were different at the $P < 0.001$ level with at least a 2-fold change between tumor and normal samples.

MDS and Difference of Gene Expression Profiles between Two Groups of Patients. MDS is one of the dimension-reduction methods used to visualize the similarity of gene expression profiles between samples (21). The similarity of the gene expression profiles between samples was assessed by Euclidean distance and Pearson correlation coefficients of the log-transformed expression ratios for the 5756 genes that were without a missing value in all 19 of the cases. The MDS plot shows sample positions in a three-dimensional space. Samples with similar gene expression profiles (short distances) are placed near each other in the MDS plot and separated from other more dissimilar groups (longer distances). When MDS plots visually appeared to show differences between groups, we formally tested these differences using the global permutation test (with 1000 permutations). Differences in individual genes were then tested using a *t* test ($P < 0.001$) for groups that were significantly different in the global permutation test.

Hierarchical Clustering. Data were evaluated using the clustering software described previously (22). Average linkage clustering in TreeView software (22) was used to generate visual representations of clusters.

Results

A total of 19 ESCC patients were analyzed, including 9 males and 10 females in this study. Six of these patients had a family history of UGI cancer, including 5 patients with cancer in a first-degree relative and 1 with cancer in a third-degree relative. Four additional patients had a family history of other cancers, and 9 had no family history of any cancer (Table 1). Clinical characteristics of patients are also shown in Table 1.

Using stringent criteria ($P < 0.001$ and ≥2-fold change) we identified 41 genes, including 13 overexpressed genes in tumors and 28 underexpressed genes (Table 2). Two-way hierarchical clustering among the 19 patients and 41 genes is shown in Fig. 1. In the underexpressed gene cluster, all but 2 of the cases (SHE1188 and SHE1190) showed a consistent pattern of expression; all of the cases in the overexpressed cluster appeared uniform. Overall, these results

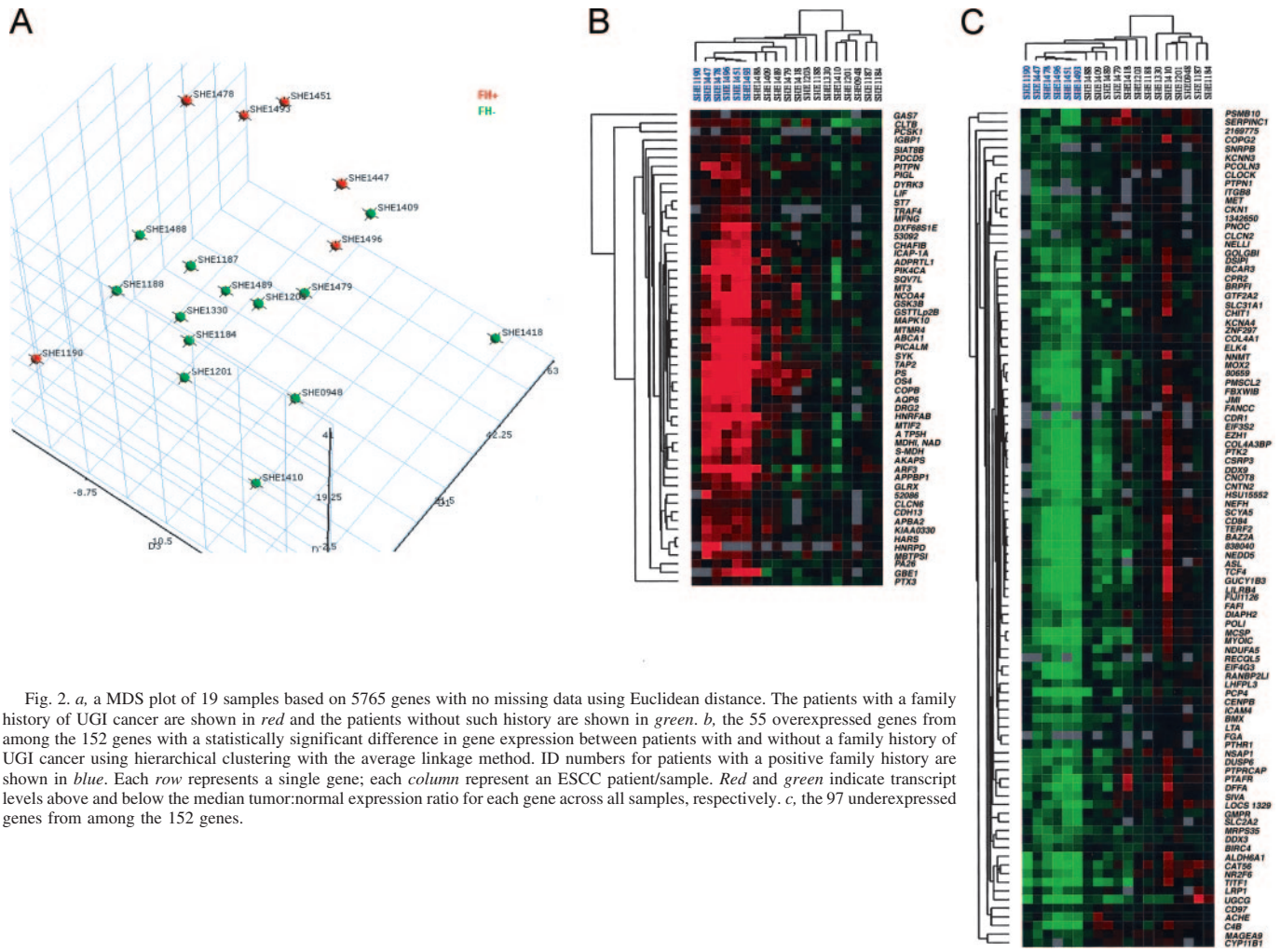

Fig. 2. *a,* a MDS plot of 19 samples based on 5765 genes with no missing data using Euclidean distance. The patients with a family history of UGI cancer are shown in *red* and the patients without such history are shown in *green*. *b,* the 55 overexpressed genes from among the 152 genes with a statistically significant difference in gene expression between patients with and without a family history of UGI cancer using hierarchical clustering with the average linkage method. ID numbers for patients with a positive family history are shown in *blue*. Each *row* represents a single gene; each *column* represent an ESCC patient/sample. *Red* and *green* indicate transcript levels above and below the median tumor:normal expression ratio for each gene across all samples, respectively. *c,* the 97 underexpressed genes from among the 152 genes.

indicate a generally consistent molecular profile of ESCC in patients from this high-risk area of China.

We examined a visual representation of relationships among all of the samples in terms of their similarities in gene expression profiles by gender, family history of cancer, and clinical characteristics with MDS using the 5765 genes with no missing values. We saw no differences in gene expression patterns by gender or clinical characteristics, but we did see clustering of cases by family history of UGI cancer (Fig. 2*a*). A global permutation test for differences in gene expression among patients with *versus* without a family history of UGI cancer was highly significant ($P = 0.007$). Finally, we identified 152 genes (includes 6 expressed sequence tags) that showed significant differences in expression between these two groups, including 55 overexpressed and 97 underexpressed genes in patients with a family history of UGI cancer ($P < 0.001$). The two-way average linkage hierarchical clustering analyses of these genes are shown in Fig. 2, *b* and *c*. As our criteria for this analysis differed from our primary analysis, only approximately half of these 152 genes had ≥1.5-fold changes, whereas just 6 had ≥2-fold changes.

Discussion

Expression array technology has become an important method for many applications, including identification of disease-related and treatment-responsive genes, and determination of carcinogenicity, toxicity, and safety of drugs (23). The goal of the current study was to exploit this technique to identify genes of which the expression correlated with ESCC because these genes could potentially be candidate molecular markers for prevention and early detection of ESCC. Hierarchical clustering of ESCC and gene expression levels showed that 13 genes were overexpressed in the tumors of most patients examined. Although some of these overexpressed genes have been described previously in ESCC [*i.e., CDC25B* (10), *COL3A1* (12), and *SPARC* (17)], for most of them this is the first report (*i.e., Fascin, LAMC2, TAGLN2, FADD*, and *CST1*). The expression alterations include genes from several categories. Of note, 5 of the 13 overexpressed genes (*COL3A1, COL7A1, SPARC, LAMC2,* and *SCYB10*) are involved in extracellular matrix functions.

Among the underexpressed genes in ESCC we identified are several genes involved in regulation of the cell cycle and proliferation (*i.e., APCL, CNN3, SPRK, FOSL1, UPK1A, EMP1, BTC,* and *PPL*), 2 genes involved in cell adhesion and cytoskeletion functions (*i.e., EVPL* and *ARC*), and 2 genes involved in the formation and maintenance of the cornified cell envelope of stratified squamous epithelia (*CSTA* and *CSTB*). More specifically, *PPL* is expressed in stratified squamous epithelia (24) and *EVPL* (25), a candidate gene for the tylosis esophageal cancer syndrome, is expressed exclusively in stratified squamous epithelia. Both *PPL* and *EVPL* have desmosome components and, thus, in conjunction with *TGM3* and *CSTA*, help to maintain an intact cell surface interface (26).

Two cases clustered differently than the other 17 (Fig. 1), particularly in regard to the underexpressed genes. One of these cases (SHE1190) was the only poorly differentiated tumor in the group. The other case (SHE1188) had the lowest percentage of tumor (50%) of the 19 specimens examined. No other differences, technical or other-

wise, were readily evident that might explain the expression differences in these two cases.

Intermediate filaments are polymers that, together with actin and microtubules, form the cytoskeleton of cells and are critical in maintaining normal cell integrity. Epithelial cell intermediate filaments are keratins derived from a family of proteins that includes *KRT14* and *KRT4*. The KRT14 protein is specifically expressed in mature ectodermal components of the esophagus. Missense mutations in *KRT14* that perturb intermediate filament assembly result in cell degeneration, disruption of the keratin network, and cell fragility. *KRT14* mutations are also associated with a group of genetic disorders known as epidermolysis bullosa simplex, a skin-fragility disorder (27). *KRT14* was first identified as an underexpressed gene in a cDNA microarray experiment using an ESCC cell line (12). However, in our study of individual tumors from ESCC patients, *KRT14* was overexpressed, ~3-fold on average. *KRT14* has also been studied as a marker of squamous differentiation in various tumors using immunohistochemistry, but only one study has reported on this technique in ESCC (28). Chu *et al.* (28) studied 435 cases of epithelial neoplasms of various primary sites of origin (including 14 ESCCs), and found that *KRT14* expression was generally restricted to squamous cell carcinomas regardless of origin and/or degree of differentiation. Thus, evidence suggests that altered *KRT14* expression may play an interesting and potentially specific role in ESCC carcinogenesis. Additional studies evaluating protein, RNA, and DNA in a larger number of cases, as well as confirmatory studies using RNA from pure tumor/normal cell populations obtained by laser capture microdissection, are needed before this potential marker can be considered for use in early detection or in the evaluation of treatment response for ESCC.

A second purpose of this study was to explore how gene expression profiles in the tumors differed by demographic and clinical characteristics. We determined that expression patterns clustered by family history of UGI cancer, and that there were 152 genes of which the expression differed significantly between persons with a positive as opposed to a negative family history of UGI cancer. We and others have been interested for many years in the potential role of genetic factors in the etiology of ESCC in high-risk areas of China. Although our previous epidemiological studies found associations between a positive family history and esophageal cancer, familial aggregation of the disease in families, and higher frequency of allelic loss in ESCC cases with a positive family history (4–6), this is the first study to provide evidence that RNA expression differs by family history status. Replication of our findings in studies with larger sample sizes is needed to confirm the association with family history.

In summary, these data indicate that molecular profiles in ESCC are highly consistent and that expression patterns in tumors from familial cases differ from those in sporadic cases. They also highlight important molecular genetic features of familial ESCC in this high-risk area of China.

Acknowledgments

We thank Drs. Ena Wang and Chandramouli Gadisetti, NCI, for technical advice.

References

1. Li, J. Y. Epidemiology of esophageal cancer in China. Natl. Cancer Inst. Monogr, *62:* 113–120, 1982.
2. Li, G., and He, L. A survey of the familial aggregation of esophageal in Yangcheng County, Shanxi Province. *In:* M Wu and D. W. Neberg (eds.), Genes and Disease. Proceedings of the First Sino-American Human Genetics Workshop, pp. 43–47. Beijing: Science Press, 1986.
3. Wu, M., Hu, N., and Wang, X. Q. Genetic factor in the etiology of esophageal cancer and the strategy of its prevention in high-incidence areas of North China. *In:* H. T. Lynch and T. Hirayama (eds.), Genetic Epidemiology for Cancer, pp. 187–200. Boca Raton, FL: CRC Press Inc., 1989.
4. Hu, N., Dawsey, S. M., Wu, M., Bonney, G. E., He, L. J., Han, X. Y., Fu, M., and Taylor, P. R. Familial aggregation of esophageal cancer in Yangcheng County, Shanxi Province, China. Int. J. Epidemiol., *21:* 877–882, 1992.
5. Hu, N., Dawsey, S. M., Wu, M., and Taylor, P. R. Family history of oesophageal cancer in Shanxi Province, China. Eur. J. Cancer, *27:* 1336, 1991.
6. Hu, N., Roth, M. J., Emmert-Buck, M. R., Tang, Z. Z., Polymeropolous, M., Wang, Q. H., Goldstein, A. M., Han, X. Y., Dawsey, S. M., Ding, T., Giffen, C., and Taylor, P. R. Allelic loss in esophageal squamous cell carcinoma patients with and without family history of upper gastrointestinal tract cancer. Clin. Cancer Res., *5:* 3476–3482, 1999.
7. Huang, J., Hu, N., Goldstein, A. M., Emmert-Buck, M. R., Tang, Z. Z., Roth, M. J., Wang, Q. H., Dawsey, S. M., Han, X. Y., Ding, T., Li, G., Giffen, C., and Taylor, P. R. High frequency allelic loss on chromosome 17p13.3-p11.1 in esophageal squamous cell carcinomas from a high incidence area in northern China. Carcinogenesis (Lond.), *21:* 2019–2026, 2000.
8. Li, G., Hu, N., Goldstein, A. M., Tang, Z. Z., Roth, M. J., Wang, Q. H., Dawsey, S. M., Han, X. Y., Ding, T., Huang, J., Giffen, C., Taylor, P. R., and Emmert-Buck, M. R. Allelic loss on chromosome bands 13q11–q13 in esophageal squamous cell carcinoma. Genes Chromosomes Cancer (Phila.), *31:* 390–397, 2001.
9. Sebastian, J. L., Rigberg, D. A., Shrivatsan, E., Revasova, E., McFadden, D. M., and Levingston, E. H. The use of large-scale cDNA analysis to profile differential gene expression in KYSE 410 human esophageal cancer cells after irradiation. Bull. Exp. Biol. Med., *130:* 882–885, 2000.
10. Hu, Y. C., Lam, K. Y., Law, S., Wong, J., and Srivastava, G. Identification of differentially expressed genes in esophageal squamous cell carcinoma (ESCC) by cDNA expression array: overexpression of Fra-1, Neogenin, Id-1, and CDC25B genes in ESCC. Clin. Cancer Res., *7:* 2213–2221, 2001.
11. Lu, J., Liu, Z., Xiong, M., Wang, Q., Wang, X., Yang, G., Zhao, L., Qiu, Z., Zhou, C., and Wu, M. Gene expression profile changes in initiation and progression of squamous cell carcinoma of esophagus. Int. J. Cancer, *91:* 288–294, 2001.
12. Kan, T., Shimada, Y., Sato, F., Maeda, M., Kawabe, A., Kaganoi, J., Itami, A., Yamasaki, S., and Imamura, M. Gene expression profiling in human esophageal cancers using cDNA microarray. Biochem. Biophys. Res. Commun., *286:* 792–801, 2001.
13. Kihara, C., Tsunoda, T., Tanaka, T., Yamana, H., Furukawa, Y., Ono, K., Kitahara, O., Zembutsu, H., Yanagawa, R., Hirata, K., Takagi, T., and Nakamura, Y. Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. Cancer Res., *61:* 6474–6479, 2001.
14. Hu, Y. C., Lam, K. Y., Law, S., Wong, J., and Srivastava, G. Profiling of differentially expressed cancer-related genes in esophageal squamous cell carcinoma (ESCC) using human cancer cDNA arrays: overexpression of oncogene MET correlates with tumor differentiation in ESCC. Clin. Cancer Res., *7:* 3519–3525, 2001.
15. Selaru, F. M., Zou, T., Xu, Y., Shustova, V., Yin, J., Mori, Y., Sato, F., Wang, S., Olaru, A., Shibata, D., Greenwald, B. D., Krasna, M. J., Abraham, J. M., and Meltzer, S. J. Global gene expression profiling in Barrett's esophagus and esophageal cancer: a comparative analysis using cDNA microarrays. Oncogene, *21:* 475–478, 2002.
16. Xu, Y., Selaru, F. M., Yin, J., Zou, T. T., Shustova, V., Mori, Y., Sato, F., Liu, T. C., Olaru, A., Wang, S., Kimos, M. C., Perry, K., Desai, K., Greenwald, B. D., Krasna, M. J., Shibata, D., Abraham, J. M., and Meltzer, S. J. Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. Cancer Res., *62:* 3493–3497, 2002.
17. Xu, S. H., Qian, L. J., Mou, H. Z., Zhu, C. H., Zhou, X. M., Liu, X. L., Chen, Y., and Bao, W. Y. Difference of gene expression profiles between esophageal carcinoma and its pericancerous epithelium by gene chip. World J. Gastroenterol., *9:* 417–422, 2003.
18. Eisen, M. B. B. P. DNA arrays of gene expression. Methods Enzymol., *303:* 179–205, 1999.
19. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T., and Marincola, F. M. High-fidelity mRNA amplification for gene profiling. Nat. Biotechnol., *18:* 457–459, 2000.
20. Huang, Y. H., Dudoit, S., Luu, P., and Speed, T. P. Normalization for cDNA microarray data. 2001. SPIE BiOS 2001. San Jose, CA, January 2001.
21. Cox, T. C. M. Multidimensional scaling, 2nd ed. Boca Raton, FL: CRC Press, 2000.
22. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA, *95:* 14863–14868, 1998.
23. Afshari, C. A., Nuwaysir, E. F., and Barrett, J. C. Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. Cancer Res., *59:* 4759–4760, 1999.
24. Aho, S., McLean, W. H., Li, K., and Uitto, J. cDNA cloning, mRNA expression, and chromosomal mapping of human and mouse periplakin genes. Genomics, *48:* 242–247, 1998.
25. Risk, J. M., Ruhrberg, C., Hennies, H., Mills, H. S., Di Colandrea, T., Evans, K. E., Ellis, A., Watt, F. M., Bishop, D. T., Spurr, N. K., Stevens, H. P., Leigh, I. M., Reis, A., Kelsell, D. P., and Field, J. K. Envoplakin, a possible candidate gene for focal NEPPK/esophageal cancer (TOC): the integration of genetic and physical maps of the TOC region on 17q25. Genomics, *59:* 234–242, 1999.
26. Steinert, P. M., and Marekov, L. N. The proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and small proline-rich proteins 1 and 2 are isodipeptide cross-linked components of the human epidermal cornified cell envelope. J. Biol. Chem., *270:* 17702–17711, 1995.
27. Chao, S. C., Yang, M. H., and Lee, S. F. Novel KRT14 mutation in a Taiwanese patient with epidermolysis bullosa simplex (Kobner type). J. Formos. Med. Assoc., *101:* 287–290, 2002.
28. Chu, P. G., Lyda, M. H., and Weiss, L. M. Cytokeratin 14 expression in epithelial neoplasms: a survey of 435 cases with emphasis on its value in differentiating squamous cell carcinomas from other epithelial tumours. Histopathology, *39:* 9–16, 2001.